

## DATA TRANSMISSION OVER A NETWORK HAVING INITIALLY UNDETERMINED TRANSMISSION CAPACITY

This invention is concerned with transmission of data from a sending station to a receiving terminal. More particularly, it envisages transmission over a telecommunications network where the transmitted bit-rate that the network can support is initially undetermined. This can arise, for example, when the rate fluctuates owing to the use of a congestion control mechanism. For example, the TCP/IP system uses IP (Internet Protocol) for transport. This is a connectionless service and simply transports transmitted packets to a destination. TCP (Transmission Control Protocol) is an overlay to this service and brings in the idea of a connection; the sending station transmits a packet and waits for an acknowledgement from the receiving terminal before transmitting another (or in the event of no acknowledgement within a timeout period, it retransmits the packet). More importantly (in the present context) it embodies a congestion control algorithm where it begins with a small packet size and progressively increases the packet size until packet loss occurs, whereupon it reduces the size again. After this initial "slow start" phase, the algorithm continues to increase the packet size (albeit more gradually) backing off whenever packet loss occurs; necessarily this involves some cyclic variation of the packet size. A description of TCP is to be found in "Computer Networks", by Andrew S. Tanenbaum, third edition, 1996, pp. 521 - 542.

Another common protocol is UDP (User Data Protocol). This does not have a congestion control mechanism of its own, but it has been proposed to add one to it, the so-called "TCP-Friendly Rate Protocol (TFRC) described in the Internet Engineering Task Force (IETF) document RFC3448. This aims to establish an average transmitting data rate similar to that which the TCP algorithm would have achieved, but with a smaller cyclic variation. It too exhibits the same "slow start" phase.

One drawback of this slow start system is that the transmitting station will not "know" what bit-rate the network will provide until the slow start phase is complete – which may take (depending on the round-trip time of the network) as much as several seconds. For some applications this does not matter, but in others it does: for example, when streaming video from a server which provides a choice of compression rates, the server cannot make an informed decision at the outset about which rate to choose. In the past, one method of dealing with this has been that the server starts by transmitting the lowest quality stream and switches up to a higher rate if and when it finds that the network can support it.

It should be stressed that the invention does not require that either of the two protocols discussed above should be used; it does however start from the assumption that one is to transmit over a connection the bit-rate of which does not become apparent until after transmission has begun.

5 According to one aspect of the present invention there is provided a method of transmitting data over a network having initially undetermined transmission capacity, in which the data comprise a first part and at least two alternative second parts corresponding to respective different resolutions, for presentation at a receiving terminal simultaneously with the first part, comprising:

- 10 (a) transmitting at least an initial portion of the first part;  
(b) receiving data indicative of the available transmission capacity;  
(c) choosing among the alternative second parts, as a function of the data indicative of the available transmission capacity;  
(d) transmitting the chosen second part and any remainder of the first part.

15

Other aspects of the invention are defined in the claims

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a transmission system in accordance with one  
20 embodiment of the invention; and

Figure 2 is a flowchart illustrating the operation of the server shown in Figure 1.

In this first example, a server 1 is to stream video and audio to a receiving terminal 2 via a network 3. Supposing that this material is pre-recorded, then the audio data is contained in a file A stored in a store 10, along with several versions V1, V2, V3 of  
25 the video, encoded at different compression rates.

At this point some explanations are in order, to avoid confusion. Reference will be made to the rate of audio or video material. This refers to the average rate at which bits were generated by the original encoder, which (apart from possible small differences in frequency references at the transmitting and receiving ends) is also equal to the  
30 average rate at which bits are consumed by the ultimate decoder. Even in constant bit-rate video compression, the instantaneous bit-rate varies according to picture content but is smoothed to a constant rate by the use of buffering. By "transmitting bit-rate" we mean the actual rate at which data are transmitted by the transmitting station.

For the purposes of the present example, we suppose that the audio file A has  
35 been encoded by means of some suitable compression algorithm at 4.8 kbit/s, whilst the

video files V1, V2, V3 are encoded at 10, 20 and 40 kbit/s respectively, perhaps using one of the well known encoding standards such as the ITU H.261, H.283, or one of the ISO MPEG algorithms.

The server 1 has a TCP interface 11, connected by a modem 12 to the network 3  
5 such as the internet. The TCP interface is entirely conventional and will not be described further. It has an input 111 for data, an output 112 for sending data packets to the modem 12, and a control output 113 which indicates to the remainder of the server whether it is permitted to deliver further data to the input 111. A control unit 13 serves to read audio and video data from the store 10, and to deliver it to the input 111 of the TCP interface 11.  
10 The data delivered to the input is also monitored by a unit 14 whose function will be described later. There is also a timer 15.

It has already been explained in the introduction that initially the server has no information about the available transmitting rate that the TCP interface 11 can deliver on to the network, and in consequence is unable to make an informed decision as to which of  
15 the three alternative video files V1, V2, V3 it should send. The rationale of the operation of this server, noting that it has only one audio file and hence no choice as to audio bit-rate, is that it delivers audio only to the interface input 111, until such time as the slow start phase of the TCP is complete (or at least, has progressed sufficiently to enable a video rate decision to be made). The purpose of the rate monitoring unit 14 is to  
20 recognise when this point has been reached. In essence, it counts the bits delivered to the interface 11, and contains (or is connected to) a timer so that it can calculate the actual transmitting bit rate that this number of bits represents. This measurement could be made over one round-trip time, but, in order to smooth out oscillations of the bit rate, one might choose to average it over a time window that is however short enough that it  
25 does not delay the recognition process significantly. Typically one might use a window length corresponding to twice (or other low multiple) of the length of the round-trip time. Thus, the monitoring unit 13 has an output 131 which indicates the current transmitting bit-rate  $R_T$ .

The system operation will now be described with reference to the flowchart  
30 shown in Figure 2. Operation begins at Step 400, where a parameter  $R_P$  representing the previous current transmitting rate is initialised to a high value, and the timer 15 is reset. At Step 401 the control unit tests the interface output 113 until it is clear to send data. Once this test is passed it reads (Step 402) audio data from the file A in the store 10 and transfers this to the interface 11. The interface 11 transmits this in accordance with  
35 normal TCP.

The control unit then interrogates the output of the monitoring unit 14 and performs some tests of the value of the current transmitting bit-rate  $R_T$  and also of the timer 15 (although naturally it is not to be expected that these tests will be passed on the first iteration). Thus if (Step 403) the transmitting rate exceeds the rate needed to transmit  
 5 audio plus full rate video (i.e. 44.8 kbit/s), further monitoring of the slow start phase becomes unnecessary and the process jumps to Step 408 (described below). If not, then at Step 404  $R_T$  is tested to determine whether it exceeds its previous value. If so it is assumed that the slow start phase is still in progress.  $R_P$  is set equal to  $R_T$  in Step 405 and the process is repeated from Step 401. If however  $R_T \leq R_P$  then the slow start phase  
 10 is deemed to have ended.  $R_P$  is set equal to  $R_T$  in Step 406 and the process moves on to a second phase. In the case of high round-trip times on the network, it can take a long time for the slow-start mechanism to conclude, and therefore also a test at Step 407 checks the state of the timer 15 and if this exceeds a maximum permitted waiting time the process moves on to the second phase where the video rate decision is then made on the  
 15 basis of the known available transmitting bit-rate, even though this might not be the maximum.

This second phase begins with the control unit making, at Step 408, a decision as to which video rate to use. In this example, it chooses the highest rate that, with the audio, represents a total bit-rate requirement less than or equal to  $R_T$ , viz.:

- 20       if  $R_T \geq 44.8$  choose V3  
           if  $44.8 > R_T \geq 24.8$  choose V2  
           if  $24.8 > R_T \geq 14.8$  choose V1  
           if  $R_T < 14.8$  5 transmission is not possible; exit at Step 409.

Once this decision has been made, the control unit then proceeds at Step 410 to  
 25 read video data from the relevant file to the TCP interface 11. It should be stressed that the initial part of this video data is contemporary (in terms of the original recorded material) with the audio already sent. Inherent in Step 410, but conventional and hence not shown explicitly, are flow control (analogous to Step 401), flow control feedback from the receiving terminal (so that the amount of data received does not cause buffer  
 30 overflow) and the possibility of switching to a higher or lower rate video file in the event that network conditions improve or deteriorate, respectively.

One issue that should be considered, though, is the fact that, because, during the start-up phase, only audio has been sent, the audio buffer at the receiving terminal is ahead of the video buffer. This may be considered desirable (to a degree at least) in  
 35 providing a greater resilience to short-term network problems for the audio than for the

video, so that in the event of packet loss causing video buffer underflow and hence loss of video at the receiving terminal, the user may continue to listen to the accompanying sound. But, if desired, the video streaming Step 410 may temporarily, during an initial period of this second phase, slow down or even suspend transmission of audio data, until  
5 the contents of the audio and video buffers at the receiving terminal reach either equality (in terms of playing time) or some other specified relationship. Naturally this has the benefit of increasing the amount of video data that can be sent during this initial period.

Possible modifications to the arrangements shown in Figure 1 include:

(a) The use of a UDP interface, with TFRC congestion control, as discussed in  
10 the introduction, in place of the TCP interface 11. In this case, because TFRC explicitly calculates the actual transmitting rate, it may be possible to dispense with the monitoring unit 13 and instead read the transmitting rate  $R_T$  directly from the UDP/TFRC interface. Recognition of the end of slow start may still be performed as shown in Step 404 of the flowchart by comparing  $R_T$  and  $R_P$ ; alternatively it may be possible to recognise it by  
15 observing when the packet loss reaches a specified level.

(b) The above description assumed that one would choose the highest video rate that the network would support; however the option also exists of deliberately choosing a lower rate in order to reduce or even eliminate the delay that occurs at the receiving terminal while the receiver video buffer is being filled to an acceptable level.  
20 Such measures are discussed in our international patent application no. PCT/GB 01/05246 (Publication no. WO 02/45372).

(c) The above description assumed that the video and audio data originated from stored files. However this method may be applied to the transmission of a live feed, provided that the server includes additional buffering so that the video can be held at the  
25 server during the initial period of audio-only transmission.

(d) Alternative audio rates can be accommodated provide a criterion can be found whereby a decision between them can be made without recourse to any information about current network conditions. An example of this might be of an internet service that can be accessed via different access routes having vastly different bandwidths, perhaps  
30 via a standard (PSTN) telephone line and a 56 kbit/s modem on the one hand and an ADSL connection at 500 kbit/s on the other. If the system has two alternative audio rates, say 4.8 kbit/s and 16 kbit/s and one makes the reasonable assumption that the PSTN connection can never support the higher rate and the ADSL connection always can, then if the server is informed by the receiving terminal (either automatically or manually) as to  
35 which type of access line is in use it can make a decision of which of the two audio rates

to choose, based on this information alone. Once that decision has been made, the process can proceed in the manner already described.

In principle, the streaming method we have described will work with a conventional receiver. However, the benefits of the proposed method will be gained only if the receiver has the characteristic that, before beginning to play the received material, it waits until both its audio buffer *and* its video buffer contain sufficient data. In general, established video standards do not specify this functionality, leaving it to the discretion of the receiver designer. Of the receivers currently available, some have this characteristic whereas others, for example, may begin to decode and play audio as soon as the audio buffer is adequately full, even when no video data have arrived. We recommend that either one chooses a receiver of the former type, or one modifies the receiver control function so as to monitor the buffer contents and to initiate playing only when both buffers contain sufficient data (in accordance with the usual criteria) to support continuous playout.

A second embodiment of the invention will now be described. This is similar to the first, except that it uses layered video coding. That is to say, instead of having several (three, in the above example) different versions of the video source only one of which is sent, one has a base layer source, which can be decoded by itself to give a low-quality video output and an enhancement layer which is useless by itself but can be decoded together with the base layer to produce a higher quality video output; and one may have further enhancement layers each of which is usable only in combination with the base layer and the intervening layer(s). In this example we also suppose that multiple (non-layered) audio rates are available. We recall that in the slow-start phase one has to transmit data in advance of deciding between the various alternative sources, and that the rationale of choosing to transmit the audio first was that since there was only one audio rate one knew that this would inevitably have to be transmitted, whatever the rate decision. In this second example with alternative audio rates this ceases to be the case, since neither audio stream qualifies as "always to be transmitted". However the video base layer does so qualify, and thus in this case one proceeds by commencing transmission of the video base layer in Step 402. Then in step 408 one selects the video and audio rates to be used and in Step 410 commences transmission of the selected audio and the enhancement layer(s), if any, appropriate to the chosen video rate. In this instance, when transmitting enhancement layer video it would be appropriate to cease transmitting base layer video until all the enhancement layer video contemporary with the base layer video already sent has been transmitted.

Of course, if only a single audio rate were used, then both audio and base layer video could be sent during the slow-start phase.

A third embodiment, similar to the second, uses video encoded using frame rate scalability in accordance with the MPEG4 standard. An encoded MPEG sequence consists of I-frames (encoded using intra-frame coding only), P-frames (encoded using inter-frame differential coding based on the content of a preceding I-frame) and B-frames (encoded using bi-directional inter-frame prediction based on neighbouring I and P-frames). A typical MPEG sequence might be IBBPBBPIBBP etc. In frame rate scaleable coding one transmits for the lowest bit-rate stream just the I-frames; for a higher bit-rate stream, the I and P-frames, and for a higher still bit-rate, all the frames. In this instance one proceeds by transmitting only I-frames at Step 402 during the slow-start phase.

A yet further example is the transmission of a page for display (a "web page") consisting of text and graphics. The idea here is slightly different from the preceding examples in that we are not now concerned with the transmission of material that has to be presented to the user in real time. Nevertheless, it is considered advantageous to provide, as we provide here, for alternative graphics resolutions. So the store 10 contains text, for example in the form of an html file, and separate image files corresponding to one or more images which the receiving terminal is to display, in conventional manner, as part of a composite display. For each image there are several, perhaps three, image files stored in the store 10, at three different resolutions. The text, or the beginning of it, is transmitted in Step 402 during the slow-start phase. At Step 408 a decision is made, based on the available transmitting rate  $R_T$ , as to which resolution to choose, the idea being that one chooses the highest resolution that allows the images to be transmitted in a reasonable period. The exit at 409 is not necessary in this case. Then at step 410 the remaining text (if any) is transmitted, followed by the file of the chosen resolution for the or each image. If the chosen files are renamed with filenames corresponding to those embedded in the text (i.e. independent of the resolution) then no modification at the receiving terminal is necessary and it can display the composite page using a standard web browser.